Volume 12 Number 19 1984 Nucleic Acids Research

Cytochrome oxidase subunit II gene of rice has an insertion sequence within the intron

Teh-hui Kao, Eunpyo Moon and Ray Wu

Section of Biochemistry, Molecular and Cell Biology, Cornell University, Ithaca, NY 14853, USA

Received 1 August 1984; Accepted 12 September 1984

ABSTRACT

We have isolated and sequenced the cytochrome oxidase subunit II gene from rice (Oryza sativa L. var Labelle). The overall structural organization of this gene is very similar to that of the maize gene. This gene contains an intron in a position identical to the intron in the maize gene. However, the intron in the rice gene is longer than that of the maize gene largely due to a 461 bp insertion sequence, which has inverted repeats at its termini and is flanked by direct repeats, characteristic of transposable elements. Apart from this insertion sequence, the remainder of the intron sequence is strikingly homologous to that of maize (98.6% homology), suggesting a possible functional or structural role. The coding regions of the two genes exhibit 99.5% nucleotide sequence homology and their deduced amino acid sequences are identical. Similarly, the 3'-noncoding regions, except for several small insertions and deletions, show complete sequence homology. On the contrary, no sequence homology is detected in the 5'-noncoding regions.

INTRODUCTION

Cytochrome oxidase, an enzyme complex found in the inner mitochondrial membrane of all aerobic organisms, catalyzes the transfer of electrons from cytochrome c to molecular oxygen[1]. The cytochrome oxidase monomer contains, along with two hemes and two coppers, at least seven distinct polypeptide subunits. Subunits I, II and III are encoded in the mitochondrial genome, whereas the other subunits are encoded in the nuclear genome. Subunit II (COII) is especially interesting because it has been shown to bind one heme and one copper atom and to be in close contact with cytochrome c. The mitochondrial COII gene has been isolated and sequenced from a variety of organisms. Animal, fungal and Drosophila COII genes do not contain introns[2-4]. Of the two plant COII genes sequenced, the maize gene contains an intron while the Oenothera berteriana gene does not [5-6]. It is interesting to note that maize is a monocotyledon and Oenothera is a dicotyledon. With additional data on a monocot presented in this report and our as yet unpublished data for a dicot, it seems reasonable to generalize

that COII genes of monocots contain introns, whereas those of dicots do not.

It has been observed that the mitochondrial genome in higher plants is considerably larger than that of animals and fungi[7]. Mitochondrial DNA also shows heterogeneity within a given plant in contrast to the uniform structure found within an animal. One goal of our present study is to determine the DNA sequence of specific mitochondrial genes, and to look for evidence of transposition, in order to better understand the dynamics of the plant mitochondrial genome.

A second goal is to identify the consensus sequence required for the splicing of introns from plant mitochondrial pre-mRNA. Determination of the intron and adjacent sequences of several mitochondrial intron-containing genes from different plants may allow us to identify the consensus signals recognized by the splicing machinery.

A third goal is to study the evolution of genes. The advantage of using the mitochondrial COII gene as a model is that it can be more easily isolated than nuclear genes because of the smaller size of the mitochondrial genome. Sequence data of the coding as well as noncoding regions of the COII gene from different plants can serve as a molecular clock to readily construct a phylogenetic tree, which can then be used to confirm and quickly expand the existing one constructed based on the protein sequence of cytochrome c[8].

MATERIALS AND METHODS

Isolation of Rice Mitochondrial DNA

Seeds of rice (Oryza sativa L., variety: Labelle) were sterilized overnight with a 1:20 dilution of Chlorox, rinsed thoroughly with sterile water, and germinated in sterile water under dark conditions in a 30°C tissue culture incubator. Ten-day-old seedlings were homogenized in a Waring blender using the following buffer: 0.4 M sucrose, 30 mM Hepes (pH 7.2), 50 mM KH_2PO_4, 5 mM EDTA, 1 mM DTT, 0.1% BSA. The homogenate was then filtered through 3 layers of Miracloth, and centrifuged at 2,000 X g for 10 min. The supernatant was centrifuged at 15,000 X g for 10 min to pellet the mitochondria. The mitochondria were gently resuspended in the homogenization buffer with a paint brush. The above low-speed and high-speed centrifugations were repeated once more. The final mitochondrial pellet was resuspended in the homogenization buffer and layered onto a discontinuous sucrose gradient (9 ml of 1 M, 8 ml of 1.3 M, and 8 ml of 1.45 M sucrose in the homogenization buffer). Centrifugation was carried out in

a SW27 rotor at 25,000 rpm for 1 hr at 4°C. Mitochondria were removed from the 1.3 M - 1.45 M interface, diluted with 3 volumes of the homogenization buffer (minus DTT and BSA), and pelleted by centrifugation at 15,000 X g for 10 min. Recovered mitochondria were lysed with 0.012% (w/v) proteinase K (Boehringer) in 50 mM Tris-HCl (pH 8), 20 mM EDTA, 2% sarkosyl at 37°C for 1 hr. Cesium chloride was added to the mitochondrial lysate at 1.17 g/ml and 200 µl of Hoechst 33258 (10 mg/ml) was added as a tracking dye. The mixture was centrifuged in a 75 Ti rotor at 65,000 rpm for 20 hrs at 15°C. The mitochondrial DNA band was withdrawn with a syringe needle and the dye removed by isopropyl alcohol extraction. The mitochondrial DNA was dialyzed against 10 mM Tris-HCl (pH 8), 1 mM EDTA.

Construction of Mitochondrial Genomic Library and Isolation of COII Clones

HindIII digested rice mitochondrial DNA was ligated to pBR322 DNA, which had been digested with HindIII and treated with calf-intestine alkaline phosphatase, at 15°C overnight. The ligated DNA was used to transform E. coli SF8 cells. Ampicillin-resistant colonies were selected and transferred to nitrocellulose filters. Colony hybridization with the nick-translated maize COII probe, a 2.4 kb EcoRI fragment from pZmE1[5], was used to screen the library for the rice COII-containing clones. Hybridizations were carried out at 23°C for 24 hrs in 50% formamide, 5 x SSC, 0.2% SDS, 5 x Denhardt's, 100 µg/ml sonicated E. coli DNA. Filters were first washed in 2 x SSC 0.2% SDS and then in 0.2 x SSC, 0.2% SDS, both at 23°C.

Restriction Enzyme Analysis

Restriction enzymes were purchased from either New England Biolabs or Bethesda Research Laboratories. Digested DNA samples were fractionated by electrophoresis in 1% agarose gels. DNA fragments were transferred to nitrocellulose filters by the method of Southern[9], and hybridized to ^{32}P-labeled probes of specific activities of 1 x 10^8 cpm/µg.

DNA Sequence Analysis

DNA sequencing was carried out by using the dideoxynucleotide chain termination procedure[10] adapted to single-stranded M13 phage[11]. A synthetic oligodeoxynucleotide primer and 80 cm polyacrylamide gels were used[12].

RESULTS AND DISCUSSION

Identification of the Rice Mitochondrial COII Gene

We used a molecularly cloned 2.4 kb maize COII gene[5] as a hybridization probe to detect the corresponding rice gene by Southern blot analysis of

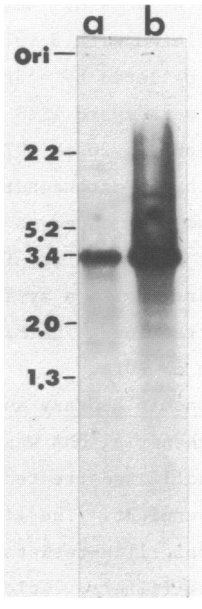

Figure 1. Southern blot analysis of HindIII-digested rice mitochondrial DNA and pOSox2. 1 µg of rice mitochondrial DNA (lane a) and 0.05 µg of pOSox2 (lane b) were digested with HindIII, electrophoresed on a 1% agarose gel, transferred to nitrocellulose and hybridized to ^{32}P-labelled nick-translated maize CO II probe (2.4 kb EcoRI insert from pZmE1, see ref. 5). Hybridization was carried out at room temperature for 36 hrs in 50% formamide, 5X SSC, 5X Denhardt solution, 0.1% SDS, 100 µg ml^{-1} of salmon sperm DNA. The filters were washed sequentially, twice in 2X SSC, 0.1% SDS and twice in 0.2X SSC, 0.1% SDS at room temperature. Size standards are in kb.

HindIII-digested rice mitochondrial DNA. As shown in Fig. 1, a 3.4 kb rice mitochondrial DNA fragment hybridized with the maize probe. A mitochondrial genomic library was constructed by ligating a HindIII digest of mitochondrial DNA to HindIII-digested and alkaline phosphatase-treated pBR322 vector. After transformation, clones containing the rice COII gene were identified by colony hybridization using the maize COII gene as a probe. The same 3.4 kb rice mitochondrial DNA insert, which hybridized to the maize COII probe in Southern blot analysis, was found in the plasmid DNA obtained from all the positive clones. The result from one of these clones is shown in Fig. 1. The plasmid contained in this clone was designated pOSox2.

A restriction map was constructed (Fig. 2) by gel electrophoretic analyses of several single or double enzyme digests of pOSox2, and the

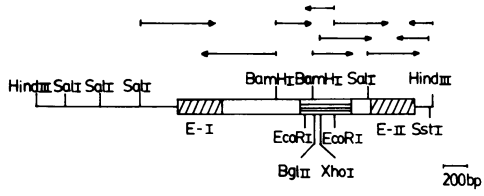

Figure 2. Restriction map of rice cytochrome oxidase subunit II gene and DNA sequencing strategy. The cross-hatched boxes represent exons, (E-I and E-II); the horizontally-lined box represents the 461 bp insertion sequence within the intron, and the open boxes indicate the remainder of the intron. The arrows above indicate the direction and extent of sequences determined from each fragment analyzed.

location of the COII gene was determined by Southern blot analysis. Those restriction fragments that contained the COII gene were then subcloned into M13 vectors mp10 or mp11 for sequence analysis.

DNA Sequence of the Rice Mitochondrial COII Gene

Sequence analysis of COII gene inserted into pOSox2 was performed by the dideoxynucleotide chain-termination method. Fig. 3 gives the sequence of the rice gene and identifies those positions where the sequence differs from that of the maize gene. The basic organization of the two genes is very similar in that there are two coding regions interrupted by a centrally located intervening sequence. Of the 783 bp in the coding sequence, only 4 bp are observed to be different, and they are all silent site substitutions. Thus, while nucleotide homology is 99.5%, homology between the deduced amino acid sequences is 100%. When the coding sequences of the COII gene from Oenothera berteriana and Pisum sativum (our unpublished results) are compared to that of the maize gene, nucleotide homology is 88.8% and 90.4%, respectively. The deduced amino acid sequence homology is 88.1% and 86.4%, respectively. These results are consistent with both the relative taxonomic similarity of the species examined as well as their relative distance on the evolutionary scale. However, the degree of sequence conservation seems much higher among the above-mentioned genes from the mitochondria of plants than among animal mitochondrial genes. For example, comparison of the COII genes of human and cow showed only 70% DNA sequence homology[13]. In contrast the COII genes of rice and maize showed 99.5% sequence homology.

Fox and Leaver[5] suggested that the codon CGG codes for Trp rather than Arg for the COII gene in maize mitochondria. Since the coding sequence for the COII gene in rice is almost identical to that in maize, we believe that CGG also codes for Trp in rice COII gene (Fig. 3, amino acid positions 57,

```
        SalI -300                                    -250                                          -200
Rice    5'-GTCGACCATTCCATAGCTTCTCGAATGGTGAGATTCAGCTTTAAGGCAGGAAGGAAATTAAAAGGAAGGAAGGAGAGTTGGGAACCCACAAAGAAAGTATCATGTTGCTCCT

                                          -150                                        -100
Rice    CGGAAAACGCGTATAGTAGTCTCATTCGCCCTTCGTCGATGGGACAAACGCTCCCAGTGTATGCGTTACAAGGCAACTAGCATTTTGCATTAGAAGTTCGTGAAAGTATCTTCT
                                                                                   Maize 5'-C CCCAAT CTCGA C  AATA

                                  -50
                                                                                -4  MET ILE LEU ARG SER LEU GLU CYS
Rice    CATTTTTCGTTGGAAAAACCAACGCCCGACGTCAAGATCAGTCTCCTTTCTAGGAGCAGACCTAAAAAAGATGGAAATTCCA ATG ATT CTT CGT TCA TTA CAA TGT
Maize   TTGG GACTACTATGTCT ATT A AAAT TATCCTTGTCTATG TAC  ACTCTCGGTTT GTCCT CTTCT  TGC G

                                  50                                                        100
        ARG PHE LEU THR ILE ALA LEU CYS ASP ALA ALA GLU PRO TRP GLN LEU GLY SER GLN ASP ALA ALA THR PRO MET MET GLN GLY ILE
Rice    CGA TTC CTC ACA ATC GCT CTT TGT GAT GCT GCG GAA CCA TGG CAA TTA GGA TCT CAA GAC GCT GCA ACA CCT ATG ATG CAA GGA ATC
Maize                                                                                   A

                                                150
        ILE ASP LEU HIS HIS ASP ILE PHE PHE PHE LEU ILE LEU ILE LEU VAL PHE VAL SER ARG MET LEU VAL ARG ALA LEU TRP HIS PHE
Rice    ATT GAC TTA CAT CAC GAT ATC TTT TTC TTC CTC ATT CTG ATT TTG GTT TTC GTA TCA CGG ATG TTG GTT CGC GCT TTA TGG CAT TTC
Maize

        200                                                         250
        ASN GLU GLN THR ASN PRO ILE PRO GLN ARG ILE VAL HIS GLY THR THR ILE GLU ILE ILE ARG THR ILE PHE PRO SER VAL ILE PRO
Rice    AAC GAG CAA ACT AAT CCA ATC CCG CAA AGG ATT GTT CAT GGA ACT ACT ATC GAA ATT ATT CGG ACC ATA TTT CCT AGT GTC ATT CCT
Maize                                                                                   T                           A

                        300                                                             350
        LEU PHE ILE ALA ILE PRO SER PHE ALA LEU LEU TYR SER MET ASP GLY VAL LEU VAL ASP PRO ALA ILE THR ILE LYS ALA ILE GLY
Rice    TTG TTC ATT GCT ATA CCA TCG TTT GCT CTG TTA TAC TCA ATG GAC GGG GTA TTA GTA GAT CCA GCC ATT ACT ATC AAA GCT ATT GGA
Maize

                                  400                                           450
        HIS GLN TRP TYR ARG SER
Rice    CAT CAA TGC TAT CGC AGT GCGCCTCTTAACGAGGGTGATTTAAGTGCAACGAAATGTACCGGTGGTTCGCCAAGCATCTGGCTTACCGGTCATCTCCCATTCCCCTCCTCC
Maize

                   500                                          550                                                 600
Rice    ACAGACTAAAAGAACTATAGCATGCCAGAAACGGAGAGTTGAGGTGGTTAGACCTATACCCCGAAATGCTCCCAGCATAGGAGCCTATGGTTCCATTCTTGTTATTGCTGGAGGTAC
Maize                                    C

                                                       650                                            700
Rice    ACATACCTCTTCTCGGTGTGGTGGAGCGATATACGAAAAATAGATGCTAAGCCTGCAATGTCCGATAACGGGGCTTTAGTAGTGAATCTATCGGCACCACAGCAGTGGCATACAACT
Maize                                                             C                    C

                                  750     BamHI                                     800
Rice    TTCGACCTAAGGCCCGGCCCCCGTTA-CTTTCGGAATGGGGGATCCCCTGTGGCAACAACCACGGTAGTAGTTGCGGAACTACTGGGCCAAGAGAGGACAACCTGTTGTTCCTGCTC
Maize                             C                    GT

                      850                                     900
Rice    CTCCTTCTTCGCTTCGGGGACGGAGG--CTACGGTAGGTAAGAGCACGCACAAGCACTTGACCGAAGGGGACCAGCGCTTCTACTCCTCCACCGAGGAGCCGTTCTTGCGAGAAGCA
Maize                               TC                           G

        950                                         1000                                                    1050
Rice    AGGGATGTCGTGAACGGTGGAGGTCAAAGAAAGAGAATTGACCTCTGAATACAGTGATCCTATGATCTAGATAGACTCCGTCCTTTTTTTTTTAGATAAGGGTGACTCAAGAAATT
Maize                                                                                                  -                   -------

                                        1100    EcoRI                                        1150          BamHI    BglII
Rice    GCGCG TGGGACCTGCTGGCATAATGCAGGCTGGAACGTGGAATTCGAGGTCTCATGAACGACTTTTTTTCGTGGACAACATACTTTCCGGTCGGCCTATGGATCCTTTTAGATCT
        ----------------------------------------------------------------------------------------------------------------

                   1200              XhoI                       1250
Rice    AGGCCCGGCCGTTCACATGAGCATAGGGAATCTATACTCGAGCCTTCGACTTGGCCCCTGACAGGATAGGTGAGGAATCACTCTTGATCTGATCTATTGGGGCCTACAACTTCGCCAA
        ----------------------------------------------------------------------------------------------------------------

        1300           1318                                  1350    EcoRI  1369                                1400
Rice    AGCCGACTAGCATCCCTTTCC ACTGCGCATTTTTCGAACAAAGAAGACTACTAAAGGATCGAATTCCCTTTCC GTGGTGAACTGGCCGTCCCATACTTCATTTTCTCTCATGTGTGT
        ----------------------------------------------------------------------------------------------------------------
                      -----------------------------------------------------

                                  1450                                             1500
Rice    TGAACATTGTCTTCCTCGGTTCCAGCTTCACTGATGTAGGTAGGGCTGGCCAGAAAGGGTCCCCTCTTGCCTATTAGTAAGCGGGCCTTCGATTCACCGGGGTCTTACGGTCTCAT
        ----------------------------------------------------------------------------------------------------------------

                  1550                                     1600                                                      SalI
Rice    AGACGCCGGAGAACTACCTAACTAAAGAAGAATAGTGCTCTTTATAAATAAGAGTAGGCGTGGAGAGCTTTTTGCGGGGAAACTTGCAAGTAAAGTTTGGGGGCAGGCGGCGTCCACC
Maize                                                                                     C

                                                                  1700
               TYR GLU TYR SER ASP TYR ASN SER SER ASP GLU GLN SER LEU THR PHE ASP SER TYR THR ILE PRO GLU ASP ASP PRO GLU LEU
Rice    CAACCT TAT GAG TAT TCG GAC TAT AAC AGT TCC GAT GAA CAG TCA CTC ACT TTT GAC AGT TAT ACG ATT CCA GAA GAT GAT CCA GAA TTG
Maize

                1750                                                             1800
        GLY GLN SER ARG LEU LEU GLU VAL ASP ASN ARG VAL VAL VAL PRO ALA LYS THR HIS LEU ARG MET ILE VAL THR PRO ALA ASP VAL
Rice    GGT CAA TCA CGT TTA TTA GAA GTT GAC AAT AGA GTG GTT GTA CCA GCC AAA ACT CAT CTA CGT ATG ATT GTA ACA CCC GCT GAT GTA
Maize

                            1850                                                            1900
        PRO HIS SER TRP ALA VAL PRO SER SER GLY VAL LYS CYS ASP ALA VAL PRO GLY ARG SER ASN LEU THR SER ILE SER VAL GLN ARG
Rice    CTT CAT AGT TGG GCT GTA CCT TCC TCA GGT GTC AAA TGT GAT GCT GTA CCT GGT CGT TCA AAT CTT ACC TCC ATC TCG GTA CAA CGA
Maize
```

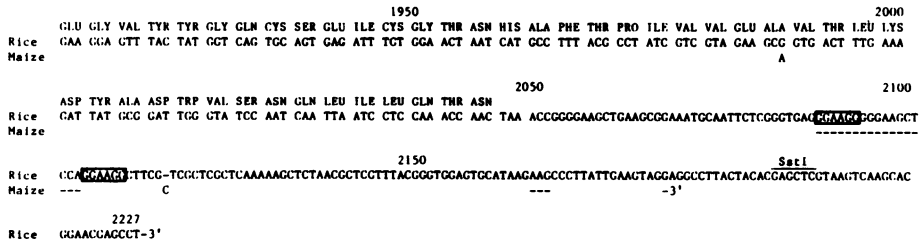

Figure 3. Nucleotide sequences of rice and maize cytochrome oxidase subunit II genes. The sequences are aligned to maximize homology except in the 5'-noncoding regions. Restriction sites are shown above the nucleotide sequence. Only those maize nucleotides which differ from rice gene are shown. Dashes indicate gaps that are necessary to align the two sequences. The vertical arrows mark the 461 bp insertion sequence in the rice intron. The boxes enclose the direct repeat sequences. Horizontal arrows indicate the inverted repeats at both ends of the 461 bp insertion sequence.

87 and 129). These suggestions are supported by the fact that at two of these three positions, the COII protein sequence in yeast, beef and Drosophila contains Trp; whereas at the third position, yeast contains Trp but beef and Drosophila lack the corresponding amino acid.

Introns and Insertion Sequences

Comparison of the introns in the COII genes of rice and maize reveals several interesting features. (1) The introns of both genes interrupt the coding sequences at precisely the same position. (2) Maize and rice introns are 794 bp and 1265 bp long, respectively. The difference in length can be accounted for largely by the presence of a 461 bp insertion sequence within the rice intron. (3) The insertion sequence is flanked on both sides by an almost perfect direct repeat of an 8 bp sequence - AGAAATTGGGGG at the 5'-junction, and AGAGGGGG at the 3'-junction as indicated in Fig. 3. Only the latter sequence is present in the maize intron. Possibly this sequence was duplicated upon the insertion of the 461 bp sequence in the rice intron. Perhaps this insertion was an early evolutionary event, and only much later was the extra 4 bp sequence AATT added into the 5'-direct repeat sequence. (4) There are imperfect inverted repeats (6 out of 7 bases) at the 5' and 3' ends of the insertion sequence, TGGGACC at the 5' end and GGTCTCA at the 3' end. It has been observed that many transposable elements begin with TG and end with CA[14]. Here, the 5' sequence does begin with TG and the 3' sequence ends with CA. This and other insertion events in the rice mitochondria suggest that they may be responsible, at least in part, for the expansion of the mitochondrial genome in plants, and provide a partial explanation why

plant mitochondrial genome is much larger than that in animals[7]. (5) Excluding the 461 bp insertion sequence and the duplicated direct repeat sequence, there are only 7 single-base substitutions, a 2 bp deletion, a 1 bp insertion and a 1 bp deletion in the remaining 794 bp intron sequence being compared. The nucleotide homology is thus 98.6%, almost as high as that of the coding sequence (99.5%).

In general, because they are not subject to functional constraints, the intron sequences of nuclear genes (exclusive of their spliced junction sequences) have evolved much faster than the coding sequences through base substitutions, deletions and insertions. This observation has been made from the comparison of genes from evolutionally related species or from multi-gene families within the same species, such as globin genes[15], cytochrome c genes[16], proinsulin genes[17], actin genes[18], and chorion genes[19]. It is surprising to find that, contrary to this general observation, the intron sequence of the COII gene, excluding the insertion sequence, has been exceedingly well conserved between rice and maize. The conservation of the intron sequence strongly suggests a possible functional or structural role, such as mRNA splicing, though additional sequence data from other plants are needed for confirmation. Michel et al. have proposed a secondary structure model for the intron sequence of the maize COII gene, in which the 5' and 3' intron - exon junctions are brought into close proximity[20]. According to this model, the 461 bp insertion sequence found within the rice intron would fit onto a 185 nucleotide hairpin loop without altering the overall secondary structure. Both rice and maize introns contain a short open reading frame of identical sequence which is continuous and in register with the first exon, terminating at the 34th codon into the intron. It is not known if this short open reading frame, together with its preceding exon, encodes a splicing enzyme, as in the case of yeast[21].

Most transposable elements in bacteria[22], yeast[23], Drosophila[24] and maize[25,26] are characterized by (1) the presence of inverted-repeat sequences at both ends of the element and (2) generation of direct-repeat sequences at the insertion sites[27]. The 461 bp insertion sequence in rice has the characteristic properties of a transposable element, namely, the duplication of an 8 bp sequence at the insertion site, and the inverted repeat sequences at the 5'- and 3'- termini. The insertion of the 461 bp sequence within the rice intron can also be viewed as a deletion of the same sequence from the maize intron. It has been proposed that this type of event occurs through slipped-mispairing of direct repeats during DNA

replication and results in the loss of the sequence between the direct repeats as well as one of the repeats itself[28]. Although we cannot distinguish between these two events, we favor the insertion event because it may explain, at least in part, the large size of plant mitochondrial DNA.

Based on the sequence data of COII genes from two monocots (rice and maize) and two dicots (Oenothera and pea), one can begin to generalize that dicots lack the intron, whereas monocots contain it. It is not known, however, whether the intron was introduced in monocots, or lost from dicots, after their divergence. This intron, similar to some of the yeast mitochondrial introns, belongs to a class of "optional" introns[21]. Borst et al. have suggested that these introns might have arisen from transposable elements which lost their mobility during the course of evolution[29]. The 461 bp insertion sequence found within the rice intron might have inserted itself into the intron through a similar mechanism. It is not yet known whether this insertion sequence is still an active transposable element or one that has already lost its mobility.

3'-Noncoding and 5'-Noncoding Sequences

Another point of interest seen in comparison of the 3'-noncoding regions of the rice and maize genes is that only insertions and deletions, but no base substitutions, are observed. As indicated in Fig. 3, there is a 17 bp insertion, a 3 bp insertion and a 1 bp deletion in the rice gene. There is a pair of 6 bp direct repeats, GGAAGG, flanking the 17 bp insertion sequence (the sequence starting at position 2088, which is underlined with dashes). Therefore, it is likely that deletion of the 17 bp sequence in maize was mediated through the above-mentioned slipped-mispairing mechanism. On the other hand, the same mechanism could result in the insertion of the 17 bp sequence in rice. Apart from these insertions and deletions, the remaining sequences (a total of 114 bp) are perfectly homologous. Blocks of homologous sequences are also observed in this region when comparing rice genes with pea genes, though the degree of homology is not as high as that in the previous comparison (unpublished results). Whether or not this sequence conservation in the 3'-noncoding region is of any functional significance remains to be tested.

In contrast to the 3'-noncoding regions, the 5'-noncoding regions of rice and maize genes show no homology at all. The DNA sequence starts to diverge at position -4 (Fig. 3). It is likely that there is an intron interrupting the 5'-noncoding region of the rice gene. The sequence complementary to the one given in Fig. 3 from positions -11 to -1 is

CCCTTTAAG$^{\downarrow}$GT, which largely conforms to the consensus sequence for an intron boundary at the 3' splice site [30]. If this is correct, the 5'-noncoding region together with any regulatory signal must reside in a location further upstream.

After these studies were completed, we learned that Bonen, Boer and Gray (personal communication) have also found an insertion sequence within the intron of the wheat COII gene at the same position which exhibits strong homology with the one that we have identified in rice. However, the insertion sequence in rice is longer, accounted for by a 52 bp long insert (positions 1318-1369 in Fig. 3). The sequence of CTTTCC present at the right-hand boundary of this insertion sequence is repeated just outside the left-hand boundary. The complement of this insertion sequence also starts with TG and ends with CA. Thus, there is an additional insertion sequence within the rice intron insert when compared to the wheat intron insert.

Acknowledgements

We thank Y.X. Cai for help in the initiation of this project, Tom Fox for the maize cytochrome oxidase subunit II gene, W. Ronnie Coffman for the rice seeds, L. Bonen, P.H. Boer and M.W. Gray for communicating their results to us, and Cindy Muncy for typing the manuscript. This work was supported by a research grant GM 29179 from the National Institutes of Health, U.S. Public Health Service.

References

1. Wikstrom, M., Krab, K. and Saraste, M. (1981) Cytochrome Oxidase - A Synthesis, Academic Press, London.
2. Fox, T.D. (1979) Proc. Natl. Acad. Sci. U.S.A. 76, 6534-6538.
3. Anderson, S., et al. (1981) Nature 290, 457-464.
4. de Bruijn, M.H.L. (1983) Nature 304, 234-241.
5. Fox, T.D. and Leaver, C.J. (1981) Cell 26, 315-323.
6. Hiesel, R. and Brennicke, A. (1983) EMBO J. 2, 2173-2178.
7. Leaver, C.J. and Gray, M.W. (1982) Annu. Rev. Plant Physiol. 33, 373-402.
8. Scogin, R. (1981) in Phytochemistry and Angiosperm Phlogeny, Young, D.A. and Seigler, D.S. Eds., pp. 19-42, Praeger Publishers, New York.
9. Southern, E.M. (1975) J. Mol. Biol. 98, 503-517.
10. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 5463-5467.
11. Messing, J. (1983) Methods in Enzymol. 101, 20-78.
12. Guo, L.H. and Wu, R. (1983) Methods in Enzymol. 100, 60-96.
13. Brown, W.M. (1983) in Evolution of Genes and Proteins, Nei, M. and Koehn, R.K., Eds., pp. 62-88, Sinauer Associates, Massachusettes.
14. Finnegan, D.J. in Genome Evolution, Dover, G.A. and Flavell, R.B., Eds., pp. 29-40, Academic Press, New York.
15. Hardies, S.C., Edgell, M.H., and Hutchison, C.A. III (1984) J. Biol. Chem. 259, 3748-3756.

16. Limbach, K.J. and Wu, R. (1983) Nucleic Acids Res. 11, 8931-8950.
17. Perler, F., Efstratiadis, A., Lomedico, P., Gilbert, W., Kolodner, R. and Dodgson, J. (1980) Cell 20, 555-566.
18. Schuler, M.A., McOsker, P. and Keller, E.B. (1983) Molec. Cell. Biol. 3, 448-456.
19. Jones, C.W. and Kafatos, F.C. (1980) Cell 22, 855-867.
20. Michel, F. and Dujun, B. (1983) EMBO J. 2, 33-38.
21. Grivell, L.A. and Borst, P. (1982) Nature 298, 703-704.
22. Starlinger, P. (1980) Plasmid 3, 241-259.
23. Cameron, J.R., Loh, E.Y. and Davis, R.W. (1979) Cell 16, 739-751.
24. Spradling, A.C. and Rubin, G.M. (1982) Science 218, 341-347.
25. Sutton, W.D., Gerlach, W.L., Schwartz, D. and Peacock, W.J. (1984) Science 223, 1265-1268.
26. Shepherd, N.S., Schwarz-Sommer, Z., Blumberg vel Spalve, J., Gupta, M., Wienand, U., Saedler, H., (1984) Nature 307, 185-187.
27. Calos, M.P. and Miller, J.H. (1980) Cell 20, 579-595.
28. Efstratiadis, A., et al. (1980) Cell 21, 653-668.
29. Borst, P. and Grivell, L.A. (1981) Nature 289, 439-440.
30. Breathnach, R. and Chambon, P. (1981) Annu. Rev. Biochem. 50, 349-384.